Image-Adaptive Watermarking Using Visual Models

Christine I. Podilchuk, *Member, IEEE*, and Wenjun Zeng

Abstract— **The huge success of the Internet allows for the transmission, wide distribution, and access of electronic data in an effortless manner. Content providers are faced with the challenge of how to protect their electronic data. This problem has generated a flurry of recent research activity in the area of digital watermarking of electronic content for copyright protection. Unlike the traditional visible watermark found on paper, the challenge here is to introduce a digital watermark that does not alter the perceived quality of the electronic content, while being extremely robust to attack. For instance, in the case of image data, editing the picture or illegal tampering should not destroy or transform the watermark into another valid signature. Equally important, the watermark should not alter the perceived visual quality of the image. From a signal processing perspective, the two basic requirements for an effective watermarking scheme, robustness and transparency, conflict with each other.**

We propose two watermarking techniques for digital images that are based on utilizing visual models which have been developed in the context of image compression. Specifically, we propose watermarking schemes where visual models are used to determine image dependent upper bounds on watermark insertion. This allows us to provide the maximum strength transparent watermark which, in turn, is extremely robust to common image processing and editing such as JPEG compression, rescaling, and cropping. We propose perceptually based watermarking schemes in two frameworks: the block-based discrete cosine transform and multiresolution wavelet framework and discuss the merits of each one. Our schemes are shown to provide very good results both in terms of image transparency and robustness.

Index Terms—**Copyright protection, DCT's, image watermarking, perceptual models, wavelets.**

I. INTRODUCTION

THE success of the Internet introduces a new set of challenging problems regarding security. One of many issues that has arisen is the problem of copyright protection of electronic information. Specifically, the idea of digital watermarking of electronic data has become an area of increased research activity over the last several years. Here we address the problem of watermarking digital image content. Current work on watermarking falls into two broad categories: source-based and destination-based schemes. Source-based schemes focus on ownership identification/authentication where a unique watermark identifying the owner is introduced to all the copies of a particular image being distributed. A source-based watermark could be used for authentication and to determine whether a received image or other electronic data has been tampered with. An important constraint to consider for many source-based applications is the ability to detect the watermark without the original image. The watermark could also be destination based where each distributed copy gets a unique watermark identifying the particular buyer or end-user. The destination-based watermark could be used to trace the end-user in the case of illegal use such as reselling. It is reasonable to assume that the content provider has the original image available for watermark detection in destination-based applications. There may be applications where we would like to attach multiple watermarks, source based as well as destination based, to one image.

We begin by reviewing some of the requirements that are necessary to provide a useful and effective watermarking scheme. These requirements apply to any data type in general but we focus on requirements that are most useful for destination-based rather than source-based applications. The three features that we examine for our application are: transparency, robustness, and capacity. Transparency refers to the perceptual quality of the data being protected. For the case of image data, the watermark should be invisible over all image types. Such a requirement is most challenging for images composed of large smooth areas. The digital watermark should also be robust to signal processing. Ideally, the amount of signal distortion necessary to remove the watermark should degrade the desired image quality to the point of becoming commercially valueless. Typical signal processing includes intentional transformations of the image data as well as illegal attempts to remove or transform the watermark into another valid watermark. Typical image transformations include compression (in particular JPEG), resampling, requantization, image enhancements, cropping, and halftoning. For destination-based watermarking, capacity may be a critical issue for widely distributed content. By capacity we mean the ability to be able to detect the watermarks with a low probability of error as the number of watermarks increases. The watermarking technique should provide a framework to insert the maximum number of distinguishable watermarks.

There has been some interesting work on source-based watermarking for authentication and alteration detection of the original data. For this application as well as several others, it is desirable to be able to extract the watermark without the original image. The requirement of being able to detect the watermark without the original image introduces a very challenging problem especially if robustness is also desirable. Here we focus on applications where the original image is available for watermark detection. Such a scheme is practical for destination-based applications such as identification of end-users (customers) where the content provider would like to identify the watermark in case of illegal use and trace the watermark back to the appropriate end-user.

Manuscript received March 1997; revised July 1997.

C. Podilchuk is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: chrisp@bell-labs.com).

W. Zeng is with Sharp Laboratories of America, Inc., Camas, WA 98607 USA (e-mail: zengw@sharplabs.com).

Publisher Item Identifier S 0733-8716(98)03272-7.

The requirements of transparency, robustness, and capacity introduce a challenging problem from the signal processing perspective. The most straightforward way to introduce a transparent watermark results in a watermark that is very vulnerable to attack. For example, placing a watermark in the least significant bits or in the high frequency components can be destroyed with simple quantization or lowpass filtering without necessarily affecting the image quality. Many of the earlier techniques used such approaches to produce visually pleasing but not robust results. A review of some of the published watermarking techniques can be found in [1].

One of the earliest papers on watermarking [2] introduced a scheme specifically for text data. The watermark consists of slight imperceivable shifts between words and lines. In the case of image watermarking, the schemes fall into two broad categories: spatial-domain and frequency-domain techniques. Spatial-domain watermarking techniques for image data include those found in [3]–[10]. The authors in [11] present an argument for why current watermarking schemes are not effective for proving ownership and propose *noninvertible* watermarking schemes to overcome this weakness. A reasonable and simple approach for ownership protection may consist of registering content in a central database. Watermark algorithms might prove to be more useful for other applications. Frequency-domain watermarking techniques could be based on spatially local or global transforms. A common transform framework for images is the block-based discrete cosine transform (DCT). One of the techniques which will be described in this paper is based on a block-DCT framework where the typical block size for the DCT is 8×8. This is the same basic decomposition currently used in the still image compression standard, JPEG. One of the first watermarking techniques based on the block-DCT is proposed in [12]. A pseudorandom subset of the blocks are chosen, and a triplet of midrange frequencies are slightly altered to encode a binary sequence. This seems to be a reasonable approach since watermarks inserted into the high frequencies are vulnerable to attack whereas the low frequency components are perceptually significant and alterations to the low frequency components may become visible. Such a scheme should provide reasonable results on average, although a more image-dependent scheme could provide better quality as well as robustness. Two DCT-based approaches are described in [13] where watermark detection does not require the original image. Another DCT-based scheme in [14] also proposes using perceptual information to embed the watermark. One significant difference between their approach and the DCT-based approach proposed here is that the visual model used in [14] results in a frequency weighting that is identical for every DCT-block in every image. In other words, the frequency weighing depends only on the basis function and does not adapt to local image characteristics or to different images. This frequency weighing is then "corrected" in the spatial domain. The DCT-based technique proposed here adapts to each DCT coefficient in every block in an optimum fashion based on threshold values determined for every DCT value in the image. These values adapt across blocks within an image as well as across images. Our DCT-based encoder is performed solely in the frequency domain in one step whereas the method proposed in [14] inserts the watermark in the spatial domain, performs frequency weighing, followed by spatial correction. Another block-based frequency domain technique described in [15] is based on inserting a watermark into the phase components of the image data since it has been established that for image data, the phase information is perceptually more significant than the magnitude data.

Several techniques have also been proposed based on global transforms of the image data. In the published literature, an interesting frequency domain method for digital watermarking of images is proposed by Cox *et al.* [1], [16] based on the idea of spread-spectrum (SS) communications. The published results show that the technique is very effective both in terms of image quality and robustness to signal processing and attempts to remove the watermark. The technique is motivated by both perceptual transparency and watermark robustness. One of the significant contributions in this work is the realization that the watermark should be inserted in the *perceptually significant* portion of the image to be robust. A DCT is performed on the whole image, and the watermark is inserted in a predetermined range of low frequency components minus the DC component. The watermark consists of a sequence of real numbers generated from a Gaussian distribution which is added to the DCT coefficients. The watermark signal is scaled according to the signal strength of the particular frequency component. This is a reasonable and simple way to introduce some type of perceptual weighing into the watermarking scheme, and the authors point out that more sophisticated models could be used. Our work is motivated by the ideas and results introduced in this paper. We propose two perceptually based schemes: 1) a block-DCT scheme which has the advantage of direct watermark encoding of JPEG bitstreams and 2) a wavelet-based scheme which provides the advantage of containing watermark components with local spatial support as well as watermark components with global spatial support, resulting in a scheme that has the benefits of both frameworks.

There has been much research over the years in trying to understand the human visual system and applying this knowledge to image processing applications. Such work has been examined for different problems with varying degrees of success. Recently, visual models have been developed specifically for the compression of image data to provide better compression than is possible using the more traditional approaches which take advantage only of signal statistics. Visual models derived for data compression are ideally suited for the digital watermarking problem. One common paradigm for perceptual coding is based on deriving an image dependent mask containing the *just noticeable differences* (JND's) which is used in compression applications to derive perceptually based quantizers and to determine perceptually based bit allocation. Such a model can be directly extended to the watermarking application by providing upper bounds on watermark intensity levels for every part of the image which guarantees transparency while providing a very robust watermark. The JND's are also useful in determining an upper bound on the number of watermarks that can be applied to a particular image

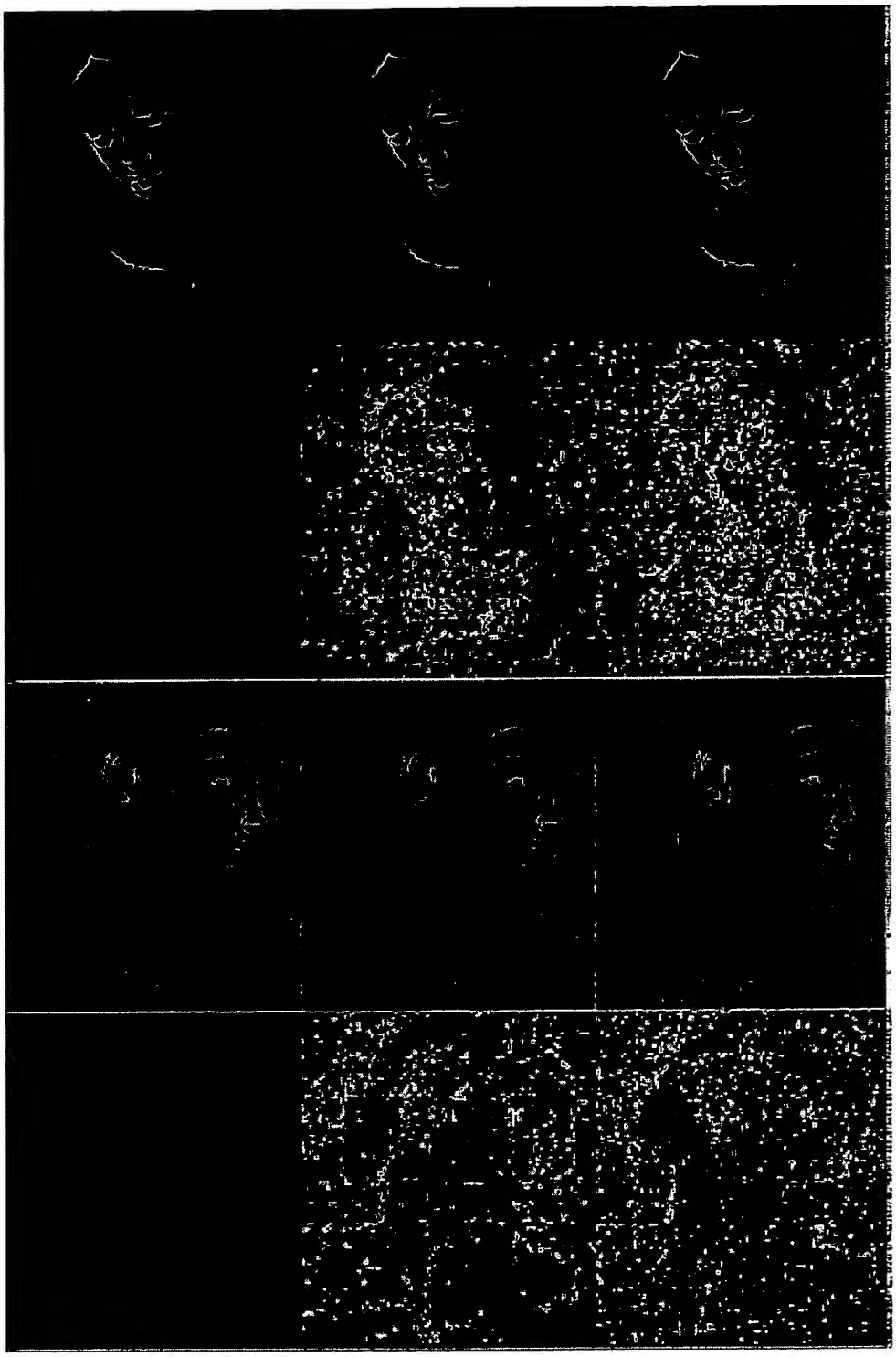

Fig. 1. Watermarked images IM1, IM2, and corresponding watermarks for SS (left), IA-DCT (center), and IA-W (right).

with low probability of error which we will refer to as the capacity problem for watermarking.

II. Motivation for Image-Adaptive Watermarking

We would like to embed a digital signature into an original image that is imperceptible and is difficult to remove without destroying the original image quality. For the receiver-based problem, we may also wish to provide the maximum number of unambiguous watermarks. We argue that by using visual models, we can adapt each watermark sequence to the local properties of the image providing a watermark that is transparent and robust. As an example, we show several images

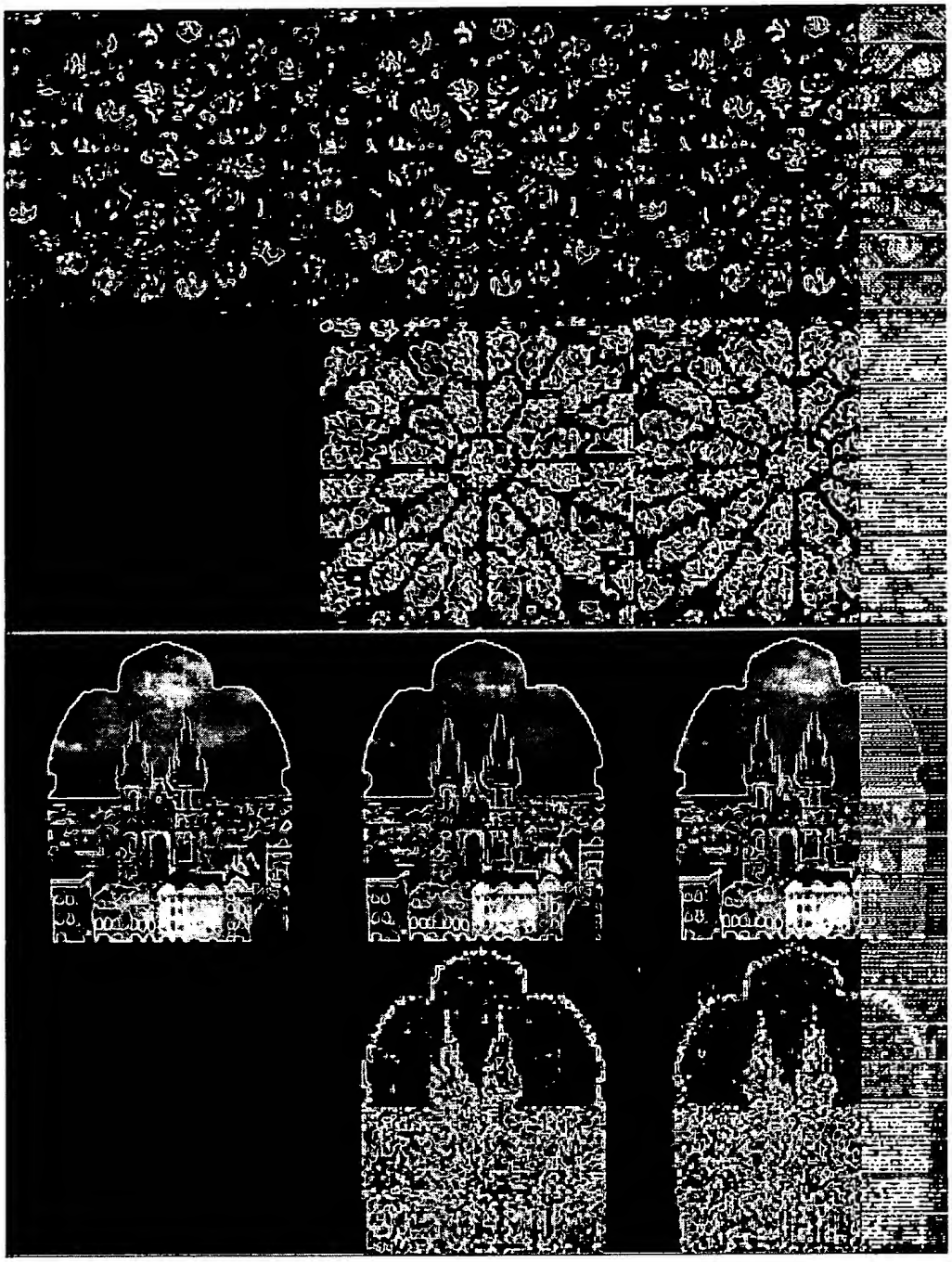

Fig. 2. Watermarked images IM3, IM4, and corresponding watermarks for SS (left), IA-DCT (center), and IA-W (right).

where image-adaptable watermarks may not provide tremendous gains due to the fairly uniform perceptual characteristics of the original images. These images are illustrated in Fig. 1. The original watermarked images are shown in the top row with the corresponding watermarks displayed in the spatial domain directly below the images for three schemes: the SS technique as described in [1], the image-adaptive DCT (IA-DCT), and image-adaptive wavelet (IA-W) schemes proposed here. Note that for these examples, the image characteristics are fairly uniform over the whole image and the perceptually based watermarks are fairly unstructured. This example is in contrast to the images shown in Fig. 2, where the image-adaptive watermarks are very structured, taking advantage of the local properties of the image. By using visual models, we can adapt the watermark to each image, providing a maximum length and maximum power watermark subject to the imperceptibility constraint. We define a watermark sequence $w_{n_I} = w_1, w_2, \cdots, w_{n_I}$ where the length of the sequence n_I is determined by our visual model and can differ for each image I. The power constraint on the watermark

TABLE I
LENGTH OF WATERMARK SEQUENCE USING THE IA-DCT WATERMARKING SCHEME

Algorithm	IM1	IM2	IM3	IM4	IM5	IM6	IM7	IM8
IA-DCT	17118	9762	60268	28679	26412	43364	70417	24936

sequence is defined as

$$\frac{1}{n_I}\sum_{i=0}^{n_I-1}(a_{I,i}w_i)^2 < P_I \tag{1}$$

where for this problem P_I is defined as the maximum power of an *imperceptible* watermark sequence and can differ for each image I. We determine both the length n_I and the weights $a_{I,i}$ using masking properties to determine a local JND $a_{I,i} = J_{I,i}$. The watermark sequence is generated from a normal distribution of zero mean and unit variance so that the power of this sequence weighted by the JND thresholds is lower than the maximum power allowed subject to the transparency condition. In other words, the original image can be thought of as the channel where the channel capacity is determined by the image characteristics. The number of unique watermarks that can be inserted into a particular image with a low probability of error can be analyzed by considering the image as a Gaussian channel where the capacity of the channel is given by $1/2\log(1 + P/N)$. For each image, the maximum power P is given by P_I which corresponds to the maximum power for a particular image subject to the transparency constraint. Likewise, for each image, the maximum length of the watermark sequence n is given by n_I which corresponds to the maximum length watermark subject to the transparency constraint. Ideally, an effective visual model should provide the maximum strength, maximum length watermark sequence that can be inserted without introducing visual distortions. This is the best we can hope to achieve in terms of capacity for a Gaussian channel, subject to perceptual transparency. Table I illustrates the length of the watermark sequence determined by the visual model used in the IA-DCT watermarking scheme described here. Note that the watermark length varies significantly depending on the particular image characteristics. The maximum allowable watermark length is simply the number of pixels in an image which for all the images considered here is 262 144 (512 × 512). This is in comparison to the method presented in [16] where the watermark sequence is a fixed length of 1000 for all images.

III. VISUAL MODELS FOR WATERMARKING

The watermarking techniques introduced here take advantage of the research results on developing useful visual models for image compression. Specifically, perceptual coders based on the JND paradigm are ideally suited in addressing the watermarking problem. For compression applications, the JND thresholds determine optimum quantization step sizes or bit allocations for different parts of the image as determined by a model of the human visual system and local image characteristics. Unfortunately, for the case of image compression, it is often the case that we cannot fully take advantage of all the masking information obtainable from a visual model. In practice, the amount of side information needed to send all the threshold values is prohibitive for image compression applications. Therefore, some average threshold values are chosen based on either the most sensitive portion of an image to provide transparent image quality at variable (usually quite high) bit rate or based on a fixed bit rate with resulting variable image quality. For instance, in JPEG compression, a visual model can be used to design one 8 × 8 quantization matrix for the whole image. This limitation does not exist for the watermarking application where we can take full advantage of the local threshold values since we have the original image at the receiver. In fact, the original image is not needed since the JND's can be quite accurately estimated from the received watermarked image. The thresholds obtained from the perceptual model are used to determine the location and maximum strength of the watermark signal that can be tolerated in every portion of the image without affecting the perceived image quality. The JND thresholds are image dependent, and as long as the watermark values remain below these thresholds, we achieve watermark transparency. As described earlier, this type of scheme also allows us to approach the maximum capacity of the given image subject to image transparency. Two things may occur with other watermarking techniques which are based on more heuristic watermarking approaches. Many times, the watermark technique is overly conservative to guarantee transparency for a wide variety of input images. A conservative approach may result in a watermark which is much weaker in some areas than a particular image could tolerate. This results in a less robust scheme than is possible by allowing the watermark signal to approach the perceptual upper bound. The length of the watermark sequence may also be conservative to avoid visible artifacts. For instance, some techniques avoid inserting a watermark signal in the low and high frequency components. The shorter sequence is less robust and lowers the watermark capacity for a given image. For some images, especially those containing large smooth areas, heuristic techniques could result in visible watermarks since the algorithms, especially those based on a global transform, are not able to adequately adapt to local image characteristics. The JND paradigm allows us to introduce a technique for adapting the watermark based on global characteristics such as viewing conditions as well as local image characteristics associated with visual masking. We describe two watermarking techniques based on visual models: an IA-DCT [17] approach and an image-adaptive wavelet IA-W approach [18].

The models used here can be described in terms of three different properties of the human visual system that have been studied in the context of image coding: frequency sensitivity, luminance sensitivity, and contrast masking. Frequency sensitivity describes the human eye's sensitivity to sine wave gratings at various frequencies. From such a model, given that the minimum viewing distance is fixed, it is possible to determine a static JND threshold for each frequency band. Frequency sensitivity provides a basic visual model which depends only on viewing conditions and is indepedent of image content. Luminance sensitivity measures the effect of

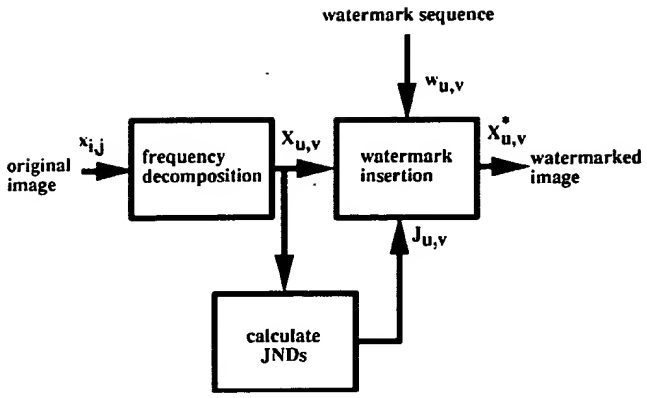

Fig. 3. Block diagram of the watermark encoder.

the detectability threshold of noise on a constant background. For the human visual system, this is a nonlinear function and depends on local image characteristics. The third component, contrast masking, allows for even more dynamic control of the JND threshold levels. Contrast masking refers to the detectability of one signal in the presence of another signal, and the effect is strongest when both signals are of the same spatial frequency, orientation, and location. Very effective visual models have been developed for compression applications that take into account frequency sensitivity, local luminance sensitivity, and contrast masking [21].

We present two watermarking schemes where the watermark insertion for both cases can be described in general as

$$X^*_{u,v} = \begin{cases} X_{u,v} + J_{u,v} w_{u,v}, & \text{if } X_{u,v} > J_{u,v} \\ X_{u,v}, & \text{otherwise} \end{cases} \quad (2)$$

where $X_{u,v}$ refers to the frequency coefficients of the original image samples $x_{i,j}$, $X^*_{u,v}$ refers to the watermarked image coefficients, $w_{u,v}$ is the sequence of watermark values generated from a normal distribution of zero mean and unit variance, and $J_{u,v}$ is the computed JND calculated for each coefficient. A block diagram of the general image-adaptive perceptual watermarking scheme is illustrated in Fig. 3. We describe two techniques: one where the frequency decomposition consists of block-based DCT's (IA-DCT scheme) and one where the frequency decomposition consists of a wavelet decomposition (IA-W scheme).

At times we may have *a priori* knowledge about some of the image transformations that will be applied to the watermarked image, and it is best to take advantage of this knowledge in the watermarking process. In this case, however, we do not assume any prior knowledge, and unlike [16] we do not limit watermark insertion only to perceptually significant parts of the image. A slight modification of (2) allows for watermark insertion to *perceptually significant* coefficients only

$$X^*_{u,v} = \begin{cases} X_{u,v} + J_{u,v} w_{u,v}, & \text{if } X_{u,v} > J_{u,v} \\ & \text{and } \dfrac{J_{u,v}}{X_{u,v}} < T_{\text{JND}} \\ X_{u,v}, & \text{otherwise} \end{cases} \quad (3)$$

where T_{JND} is an empirically derived threshold value that determines the cutoff for perceptually significant frequency components as determined by the JND threshold values.

A. Image-Adaptive DCT (IA-DCT) Watermarking

JPEG is the current international standard for color still image compression. Therefore, it is important to examine how we can take advantage of visual models within this framework even though block-based DCT's are not ideal in terms of mimicking our visual system's structure. For details on JPEG compression see [19]. For the watermarking scheme using the DCT framework, the original image is decomposed into nonoverlapping 8 × 8 blocks and the DCT is performed independently for every block of data. Due to the block-structure in the decomposition, we will refer to the original image pixels as $x_{i,j,b}$ where i, j denotes the location in block b and $X_{u,v,b}$ denotes the DCT coefficient for the basis function associated with location u, v in block b. Since JPEG allows the user to specify a quantization table for each image, it should be possible to derive a "perceptually optimal" one. In [20] a set of formulas for determining the perceptually optimal quantization matrices for both luminance and chroma given the viewing conditions is presented. This model takes into account frequency sensitivity in determining the optimum quantization matrix but does not take into account the image dependent components of luminance sensitivity and contrast masking. This has been addressed by Watson [21], where this approach has been extended to determine an image dependent quantization table that incorporates not only the global conditions, but also accounts for local luminance and contrast masking. These thresholds are used to derive an optimal image dependent quantization table for a specified level of visual distortion. Since JPEG allows for only one quantization matrix for all the image blocks, it is difficult to take full advantage of the local properties as given by Watson's model. The work in [22] introduces additional local quantizer control by using Watson's model to drive a prequantizer which zeros out all DCT coefficients below the locally derived JND threshold but also cannot utilize all the local threshold information. The JPEG bitstream specification limits the amount of perceptual fine tuning that can be incorporated into the coder. This applies to other coding schemes as well where much of the information obtained from an image-dependent visual mask cannot be incorporated into the design of a coder due to the amount of overhead needed to transmit the side information. For the watermarking application, however, we are not limited by the amount of bits needed to transmit the perceptual mask since the original image is available at the receiver for watermark detection. The JND thresholds are directly calculated from the original image.

The 8 × 8 DCT framework provides some local control which allows us to incorporate local visual masking effects into the watermark encoder although such a decomposition is not ideally suited for taking advantage of visual masking. A benefit of such a scheme is that if the images are stored as compressed JPEG bitstreams, the watermarks can be inserted directly into the bitstream by partially decompressing the data. This is in contrast to decoding the image, applying the watermark, and encoding the image again.

For the watermarking problem, we can fully utilize the local information extracted from the visual models since the original

image is available at the receiver. We originally examined two perceptual models which have been applied to the baseline mode of the JPEG coder [21], [23]. Both the Watson model [21] and the Safranek–Johnston model [23] are based on the same image independent component utilizing frequency sensitivity as determined by measurements of specific viewing conditions. This component of the visual model is based on the work presented in [20] with a minimum viewing distance of four picture heights and a D65 monitor white point. We refer to the frequency sensitivity portion of the model as $t^F_{u,v}$ where a frequency threshold value is derived for each DCT basis function and in this case results in an 8 × 8 matrix of threshold values. Watson further refines this model by adding a luminance sensitivity and contrast masking component [21]. Luminance sensitivity is estimated by the formula

$$t^L_{u,v,b} = t^F_{u,v}\left(\frac{X_{0,0,b}}{\bar{X}_{0,0}}\right)^a \tag{4}$$

where $X_{0,0,b}$ is the DC coefficient of the DCT for block b, $\bar{X}_{0,0}$ is the DC coefficient corresponding to the mean luminance of the display, and a is a parameter which controls the degree of luminance sensitivity. The authors in [20] suggest setting a to 0.649. Given a DCT coefficient $X_{u,v,b}$ and a corresponding threshold value derived from the viewing conditions and local luminance masking, $t^L_{u,v,b}$, a contrast masking threshold $t^C_{u,v,b}$ is derived as

$$t^C_{u,v,b} = \text{Max}\left[t^L_{u,v,b}, |X_{u,v,b}|^{w_{u,v}}\left(t^L_{u,v,b}\right)^{1-w_{u,v}}\right] \tag{5}$$

where $w_{u,v}$ is a number between zero and one and can assume a different value for each DCT basis function. A typical empirically derived value for $w_{u,v}$ is 0.7.

The Safranek–Johnston model [23] is composed of the same image independent component $t^F_{u,v}$ as Watson's model with a different contrast masking component. The model for contrast masking is a function of the standard deviation of the DCT coefficients and is described as

$$t_{C,b} = \begin{cases} 1.0, & \text{if } \sigma_b < t_{\min} \\ 1.0 + \dfrac{\sigma_b - t_{\min}}{(t_{\max} - t_{\min})(T_{\max} - 1)}, & \text{if } t_{\min} \le \sigma_b \le t_{\max} \\ T_{\max}, & \text{if } \sigma_b > t_{\max} \end{cases} \tag{6}$$

where $t_{\min}$ and $t_{\max}$ are empirically derived lower and upper threshold values, σ_b is the standard deviation of the DCT coefficients in block b, and $T_{\max}$ is an empirically derived maximum elevation level. For more details on the derivation of this model for contrast masking, please refer to [24]. The Safranek–Johnston model presents a very simple technique for estimating contrast masking but results in only one threshold value for an entire block. This technique does not take into account the distribution of energy within the block and cannot distinguish between a block with an edge or some other type of structured high frequency content and a block of random texture. Perhaps for these reasons, when comparing the watermarking scheme using Watson's model and the Safranek–Johnston model, Watson's model resulted in better image quality. This is consistent with the published results presented in [22] where the author shows that for image compression, the results are similar. For the rest of the discussion here, as well as the section on results, the IA-DCT algorithm assumes Watson's model for determining the JND's.

The image dependent masking thresholds are used to determine the location and maximum strength of the watermark which consists of a sequence of real numbers generated from a Gaussian distribution with zero mean and unit variance as proposed in the SS technique in [16]. One reason for such a watermark is robustness to collusion and is described by the authors in [1].

The watermark encoder for the IA-DCT scheme is described as

$$X^*_{u,v,b} = \begin{cases} X_{u,v,b} + t^C_{u,v,b} w_{u,v,b}, & \text{if } X_{u,v,b} > t^C_{u,v,b} \\ X_{u,v,b}, & \text{otherwise} \end{cases} \tag{7}$$

where $X_{u,v,b}$ refers to the DCT coefficients, $X^*_{u,v,b}$ refers to the watermarked DCT coefficients, $w_{u,v,b}$ is the sequence of watermark values, and $t^C_{u,v,b}$ is the computed JND calculated from the visual model described in [21]. Note that since the watermark is generated from a normal distribution, watermark insertion as given in (2) will occasionally result in values that exceed the JND. Informal studies show that exceeding the JND occasionally does not result in any visibly objectionable results. This might signify that there are other masking effects that could be incorporated into the visual models that we are not currently taking advantage of. We have not run formal tests, however, to make any definite conclusions. The DCT-based technique offers the advantage of direct watermarking of JPEG bitstreams by partially decompressing the bitstream. Specifically, the bitstream is passed through the entropy decoder and inverse quantizer and then watermarked. Currently, the watermark is only inserted into the luminance component of the image.

B. Image-Adaptive Wavelet (IA-W) Watermarking

A perceptually based watermarking algorithm is also proposed based on a wavelet decomposition where the threshold values have also been derived previously for image compression [25]. Frequency sensitivity thresholds are determined for a hierarchical decomposition using the 9-7 biorthogonal filters from [26]. Due to the hierarchical decomposition, this approach has the advantage of constructing watermark components that have varying spatial support providing the benefits of both a spatially local and a spatially global watermark. The watermark component with local spatial support is suited for local visual masking effects and is robust to signal processing such as cropping. The watermark component with global spatial support is robust to operations such as lowpass filtering. Due to the hierarchical nature of such an approach, this scheme is more robust to certain types of distortions than the DCT-based framework which produces watermarks with only local spatial support or to the SS approach which produces watermarks with only global spatial support. The wavelet framework consists of a four-level decomposition as illustrated

in Fig. 5. Here, the upper left-hand corner corresponds to the lowest frequency band. The visual model used here is much simpler than the one used in the DCT-based scheme. A weight $t^F_{l,f}$ is determined for each frequency band based on typical viewing conditions. Here l denotes the resolution level where $l = 1, 2, 3, 4$ and f denotes the frequency orientation where $f = 1, 2, 3$. Referring to Fig. 5 the frequency locations 1 and 2 refer to low horizontal/high vertical frequency components and low vertical/high horizontal frequency components, respectively, and frequency location 4 refers to high horizontal/high vertical frequency components. The details of the experiments and resulting weights can be found in [25]. This model could be further refined by adding image-dependent components as in the DCT-based approach. Even this simple visual model, however, yields very good results, and the hierarchical framework provides a robust watermark as well as finer control of watermark insertion than can be obtained using a block-based scheme. Results comparing the wavelet-based scheme to the DCT-based scheme will be described in the section on results. The watermark insertion for IA-W is described by

$$X^*_{u,v,l,f} = \begin{cases} X_{u,v,l,f} + t^F_{l,f} w_{u,v,l,f}, & \text{if } X_{u,v,l,f} > t^F_{l,f} \\ X_{u,v,l,f}, & \text{otherwise} \end{cases} \tag{8}$$

where $X_{u,v,l,f}$ refers to the wavelet coefficient at position (u, v) in resolution level l and frequency orientation f, $X^*_{u,v,l,f}$ refers to the watermarked wavelet coefficient, $w_{u,v,l,f}$ is the watermark sequence, and $t^F_{l,f}$ corresponds to the computed frequency weight at level l and frequency orientation f for the 9-7 biorthogonal filters. As for the IA-DCT approach, the watermark is inserted only in the luminance component of the image.

IV. Watermark Detection

Watermark detection is based on classical detection theory [27], and a block diagram of the process appears in Fig. 4. This is the same basic approach used in the SS detection scheme [16]. The original image is subtracted from the received image, and the correlation between the signal difference and a specific watermark sequence is determined. The correlation value is compared to a threshold to determine whether the received image contains the watermark in question. The normalized correlation detection scheme for the IA-DCT scheme can be expressed as

$$w^*_{s,u,v,b} = X_{u,v,b} - \hat{X}^*_{u,v,b}$$
$$w^*_{u,v,b} = \frac{w^*_{s,u,v,b}}{t^C_{u,v,b}} \tag{9}$$

$$\rho_{ww^*} = \frac{w^* \cdot w}{\sqrt{E_w E_w}} \tag{10}$$

where $w^* \cdot w$ denotes the dot product, $w^*_{s,u,v,b}$ denotes the possible received, perhaps distorted watermark scaled by the JND thresholds $t^C_{u,v,b}$, $w^*_{u,v,b}$ denotes the received watermark, and ρ_{ww^*} is the normalized correlation coefficient between the two signals w and w^*. If w is identical to w^* and normally

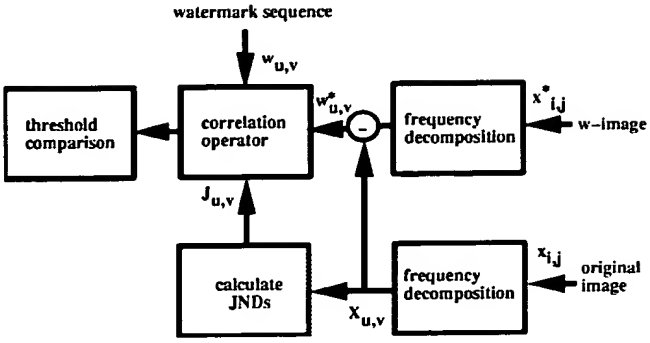

Fig. 4. Block diagram of the watermark decoder.

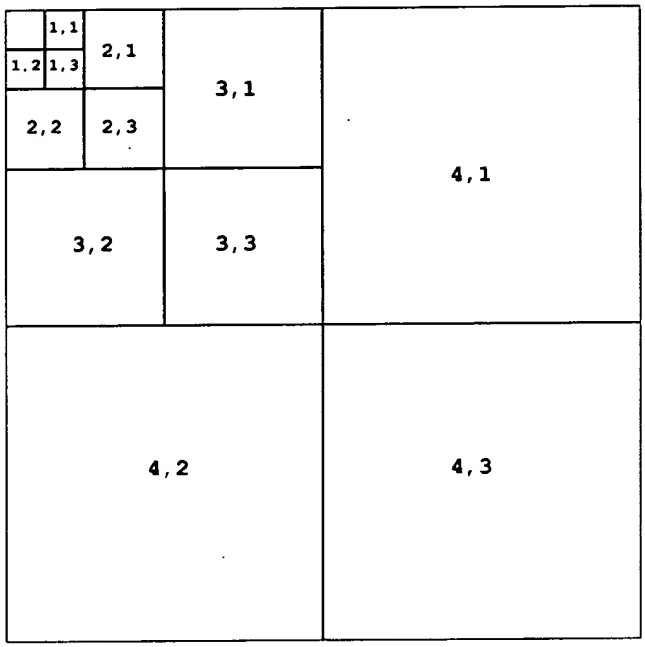

Fig. 5. Wavelet decomposition for IA-W scheme.

distributed, the correlation coefficient is one. If w and w^* are independent, ρ_{ww^*} is also normally distributed. Therefore, the probability of ρ_{ww^*} exceeding a certain threshold can be calculated from the normal distribution. The watermark detection is performed by comparing the correlation coefficient to a threshold value which can be modified according to the tradeoff between probability of detection and the probability of false alarm that is appropriate for a particular application. The final step for watermark detection is

$$\begin{aligned} \rho_{ww^*} &> T_\rho \quad \text{watermark } w \text{ detected} \\ \rho_{ww^*} &\le T_\rho \quad \text{watermark } w \text{ is not detected.} \end{aligned} \tag{11}$$

Any prior knowledge about the image transformations should be incorporated either in the watermark encoder or decoder. For instance, if it is known that the image is to be lowpass filtered in some way, the high frequency components should be avoided for watermarking. At the decoder, we perform some postprocessing on the received image to "whiten" our received sequence to achieve better detection results. The work presented by [16] offers several techniques to estimate the degradations of the received,

possibly watermarked image based on the original image. We employ several similar techniques such as using the original image to register a cropped image plus watermark. We have also employed several simple steps to ignore any received watermark values that deviate from what is expected. This filtering procedure is applied to the detection step of all the watermarking techniques we compare. The filtering step is described in general for any frequency location (u, v) as

$$W^*_{u,v} = \begin{cases} 0, & \text{if } \hat{X}^*_{u,v} < J_{u,v} \\ w^*_{u,v}, & \text{otherwise} \end{cases} \tag{12}$$

where $J_{u,v}$ is determined by the JND's for the image-adaptive schemes and by a scaled version of the frequency coefficient for the SS scheme. We discard any watermark signals corresponding to frequency locations where the received watermarked signal has fallen below a threshold.

For the IA-W algorithm, the correlation is performed separately at each level as labeled in Fig. 5, that is

$$\begin{aligned} w^*_{s,u,v,l,f} &= X_{u,v,l,f} - \hat{X}^*_{u,v,l,f} \\ w^*_{u,v,l,f} &= \frac{w^*_{s,u,v,l,f}}{t^F_{l,f}} \end{aligned} \tag{13}$$

$$\rho_{ww^*}(l, f) = \frac{w^*_{l,f} \cdot w_{l,f}}{\sqrt{E_{w_{l,f}} E_{w_{l,f}}}} \quad \text{for } l = 1, 2, 3, 4 \text{ and } f = 1, 2, 3. \tag{14}$$

In this case the normalized correlation is calculated separately for each subband (l, f). We calculate the average for each resolution level l as

$$\rho_{ww^*}(l) = \frac{1}{N_f} \sum_{f=1}^{N_f} \rho_{ww^*}(l, f) \qquad \text{for } l = 1, \cdots, 4 \tag{15}$$

where N_f is the number of frequency orientations. In this case $N_f = 3$. By evaluating the correlations separately at each resolution, we can use this to our advantage in the detection process. For instance, cropping the image will impact the watermark values in the lower frequency levels more than in the higher levels. This is due to the fact that the watermark sequence in the higher levels corresponds to a smaller spatial support. Likewise, any type of lowpass filtering operation will affect the higher frequency watermark coefficients more than the lower levels. We can take advantage of this by discarding layers with low correlation values. Similarly, we calculate the average correlation value over a certain frequency orientation, i.e.,

$$\rho_{ww^*}(f) = \frac{1}{N_l} \sum_{l=1}^{N_l} \rho_{ww^*}(l, f) \qquad \text{for } f = 1, \cdots, 3 \tag{16}$$

where N_l is the number of levels. In this case $N_l = 4$. By evaluating the correlations separately for each frequency orientation, we can take advantage of any strong structure that is associated with the original image where the watermark sequence is much stronger than in other frequency orientations. By examining the subband correlations separately, we can choose the maximum correlation value over all the possible levels as well as frequency locations

$$\rho^*_{ww^*} = \max_{l,f} \{\rho_{ww^*}(l), \rho_{ww^*}(f)\}. \tag{17}$$

We now compare our image-adaptive perceptual algorithms IA-DCT and IA-W to the SS technique outlined in [1], to see if there are any gains in using image-adaptive watermarking based on a formal visual model over a less image-dependent scheme.

V. Results

We apply the perceptual watermarks using the DCT framework and the wavelet framework to a wide variety of images to test both for transparency and robustness. Since the perceptual watermark schemes adapt on a local level to image characteristics, we hope to get better results than techniques which do not adapt to local characteristics. The SS technique described in [16] presents a wide range of robustness results without adapting to local image properties. We compare our technique to the SS technique to see if there are any gains in using formal visual models for image-adaptive watermarking. Our implementation of the SS technique is consistent with the results reported in [16]. Our detection scheme is consistent for all the techniques we tested as outlined in the previous section. It is possible that other postprocessing techniques may help yield better detection results. In our comparisons, we examine detectability of one watermark. We extend the framework here to the problem of detectability of multiple watermarks and the capacity issue in the work to be presented in [28]. It is also worth mentioning that, although the watermark detection scheme used here assumes that the original image is available at the decoder, the general framework could be modified for detection without the original image. The basic difference is that the JND values will be estimated from the received watermarked image and the correlation values will drop resulting in a less robust scheme.

Image Quality: We examined a wide range of pictures for image quality. It is reasonable to assume that many of the published watermarking techniques will yield satisfactory results in terms of transparency for images with relatively high complexity, that is images which contain a fair amount of details and texture. The transparency of watermarks in images with high frequency content is actually an example of contrast masking. The images illustrated in Figs. 1, 2, and 6 show some typical images which yield good results in terms of transparency for the watermarking techniques considered here. It is likely that any conservative watermark encoding scheme based on some simple, heuristic rules will provide acceptable image quality for a wide range of image types. In comparing the perceptual methods described here and the SS technique in [16], differences in image quality become apparent for images that contain large smooth areas. For such images, the SS watermark may become visible. Fig. 7 illustrates two images where the SS watermark affects the original image quality. The two images shown illustrate the watermarked images using the SS algorithm on the left, the IA-DCT algorithm in the center, and IA-W algorithm

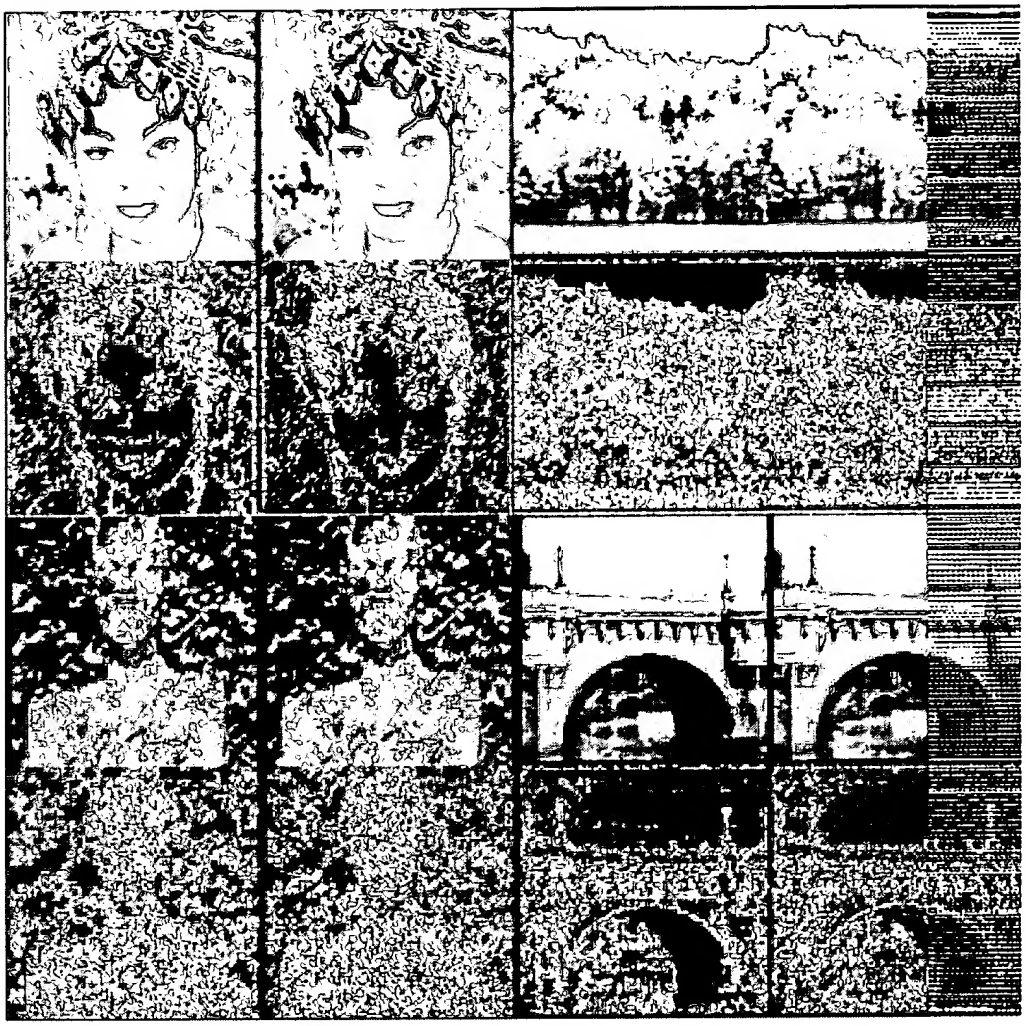

Fig. 6. Watermarked images IM5–IM8 and corresponding watermarks for IA-DCT (left) and IA-W (right).

on the right with the corresponding watermarks illustrated below each image. The SS watermark is most visible in the smooth background area where the "blotchy" distortion has a similar structure to the watermark. Note how the image-adaptive approaches avoid inserting strong watermark signals in the large smooth areas of the picture. This is especially pronounced in the "skier" (IM9) where the perceptually based watermarks are guided to the edges of the image data. Note that adjusting the scaling parameter to a lower value for the SS approach, from the original vale of 0.1 given in [1], results in a less visible watermark at the expense of robustness. Even better results could be obtained if the scaling parameter is adapted to each frequency component in some perceptual manner, although in the current framework, local spatial control of the watermark is not possible. We compare the proposed watermarking schemes to the SS approach using several images of varying characteristics. IM1–4 are illustrated in Figs. 1 and 2, and IM5–8 with the corresponding IA-DCT and IA-W watermarks are illustrated in Fig. 6.

Robustness to JPEG Compression and Cropping: We examine watermark robustness to JPEG compression and cropping. We can think of these as approximately dual problems where cropping zeros out spatial components and JPEG compression (or more generally any type of lowpass filtering operation) zeros out frequency components. Since the IA-DCT watermark framework is the same as for JPEG compression, we expect the SS as well as the IA-W schemes to be more robust to JPEG compression. Also due to the fact that the SS approach avoids inserting watermark components into the high frequencies, we expect such a technique to be more robust to compression as well as any general lowpass filtering operation. Likewise, we expect the schemes which introduce watermarks with local spatial support, namely IA-DCT and IA-W, to have an advantage over the SS technique for spatial cropping.

Table II shows the correlation results after JPEG compression. Each column refers to a different quality factor Q where lower Q values correspond to greater compression. For most images, the image quality suffers significantly for Q values lower than 40, resulting in very visible blocking artifacts. We compare the SS approach to IA-DCT and IA-W for several different images shown in Figs. 1, 2, and 6. It is interesting to observe that intuitively IM1 and IM2 as illustrated in Fig. 1 seem to have little to gain by image-adaptive watermarking as illustrated by their relatively "unstructured" watermarks. The

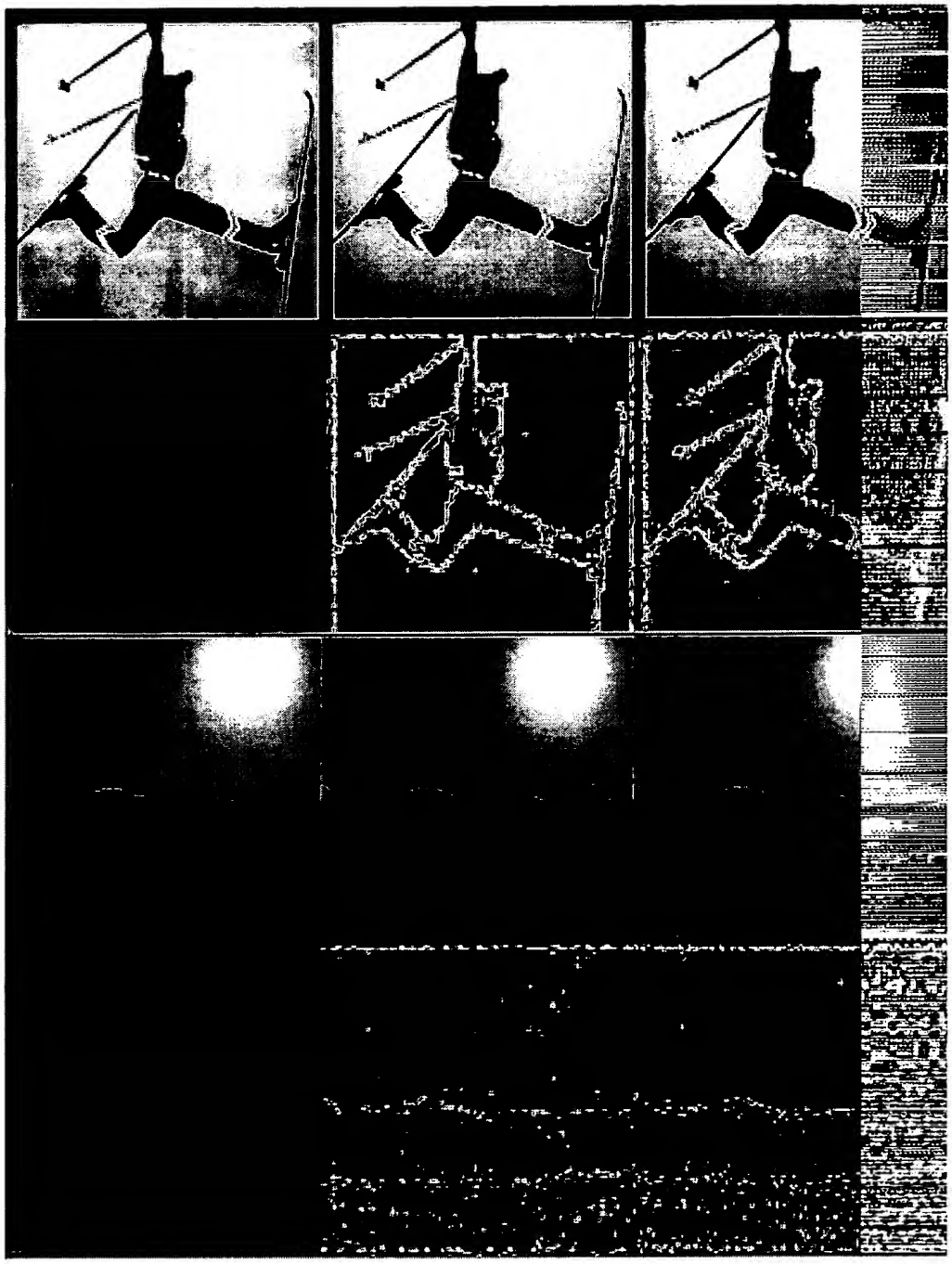

Fig. 7. Watermarked images IM9, IM10, and corresponding watermarks for SS (left), IA-DCT (center), and IA-W (right).

detection results using IA-DCT and IA-W, however, are consistently better than the SS approach for these images. In other words, there are gains in using image-adaptive watermarking even for images with seemingly uniform perceptual characteristics. For IM3, the IA-W approach yields the best results, followed by the SS approach and the IA-DCT approach. In general, for most images, all three techniques are robust to JPEG compression with the IA-W scheme outperforming IA-DCT and SS. Since the IA-DCT scheme is inserted in the same framework as JPEG compression, it would seem that such a scheme would not be robust to JPEG compression. This is not the case, however, from the results presented here. Table III illustrates the detector outputs for JPEG compression for two images consisting of large smooth areas which are difficult to watermark in a transparent fashion. These images are illustrated in Fig. 7 where the SS watermark results in a visible streaking effect noticeable in the large smooth areas of the picture. We compare the output of the SS approach using the specified scaling factor of 0.1 as described in [16] as well as a modified version of the SS approach where the scaling factor is lowered until the watermark is no longer visible to compare results of the different schemes when the

TABLE II
COMPARISON OF ROBUSTNESS TO JPEG COMPRESSION

Image	Algorithm	quality factor Q					
		80	60	40	20	10	5
IM1	SS	0.57	0.43	0.48	0.35	0.2	0.08
	IA-DCT	0.90	0.75	0.62	0.48	0.36	0.25
	IA-W	0.93	0.89	0.79	0.75	0.46	0.18
IM2	SS	0.66	0.55	0.46	0.21	0.15	0.07
	IA-DCT	0.91	0.81	0.68	0.52	0.4	0.23
	IA-W	0.97	0.91	0.82	0.67	0.55	0.32
IM3	SS	0.9	0.8	0.68	0.53	0.43	0.23
	IA-DCT	0.8	0.66	0.56	0.42	0.29	0.16
	IA-W	1.0	0.95	0.75	0.54	0.45	0.44
IM4	SS	0.86	0.76	0.62	0.67	0.46	0.22
	IA-DCT	0.9	0.77	0.66	0.51	0.38	0.24
	IA-W	0.98	0.88	0.81	0.68	0.42	0.2
IM5	SS	0.95	0.94	0.9	0.88	0.7	0.42
	IA-DCT	0.9	0.8	0.69	0.54	0.39	0.24
	IA-W	1.0	0.95	0.9	0.84	0.68	0.48
IM6	SS	0.44	0.57	0.34	0.27	0.13	0.13
	IA-DCT	0.87	0.74	0.62	0.47	0.35	0.24
	IA-W	0.93	0.83	0.75	0.86	0.35	0.31
IM7	SS	0.94	0.71	0.72	0.61	0.5	0.3
	IA-DCT	0.94	0.79	0.67	0.52	0.36	0.21
	IA-W	0.98	0.94	0.89	0.71	0.55	0.27
IM8	SS	0.91	0.85	0.85	0.67	0.49	0.31
	IA-DCT	0.9	0.78	0.66	0.51	0.38	0.22
	IA-W	0.99	0.9	0.86	0.62	0.39	0.19

TABLE III
COMPARISON OF ROBUSTNESS TO JPEG COMPRESSION

Image	Algorithm	quality factor Q					
		80	60	40	20	10	5
IM9	SS (visible)	0.87	0.84	0.80	0.65	0.49	0.25
	SS (invisible)	0.37	0.25	0.23	0.14	0.08	0.04
	IA-DCT	0.84	0.73	0.64	0.49	0.33	0.18
	IA-W	0.95	0.92	0.74	0.62	0.62	0.39
IM10	SS (visible)	0.26	0.17	0.13	0.09	0.04	0.02
	SS (invisible)	0.13	0.01	0.0	0.01	0.05	0.03
	IA-DCT	0.87	0.76	0.66	0.50	0.35	0.12
	IA-W	0.95	0.93	0.93	0.76	0.24	0.25

watermarks are all transparent. In this case, the scaling factor was reduced to 0.025 and 0.035, respectively, to produce transparent watermarks for the SS approach. Again, for all cases, the IA-W scheme yields the best results on average followed by the IA-DCT approach and SS approach.

Table IV shows correlation results after cropping the original image to one-quarter of its original size followed by JPEG compression. Table V shows similar results after cropping

TABLE IV
COMPARISON OF ROBUSTNESS TO JPEG COMPRESSION + CROPPING (1/4 ORIGINAL)

Image	Algorithm	quality factor Q				
		crop	crop+Q80	crop+Q60	crop+Q40	crop+Q20
IM1	SS	.27	.23	0.23	0.21	0.14
	IA-DCT	1.0	0.89	0.77	0.63	0.48
	IA-W	1.0	0.97	0.92	0.90	0.89
IM2	SS	0.12	0.12	0.11	0.13	0.11
	IA-DCT	1.0	0.9	0.82	0.67	0.52
	IA-W	1.0	1.0	0.97	0.9	0.9
IM3	SS	1.0	0.7	0.8	0.7	0.63
	IA-DCT	1.0	0.8	0.68	0.57	0.44
	IA-W	1.0	0.98	0.91	0.89	0.86
IM4	SS	0.57	0.58	0.58	0.48	0.48
	IA-DCT	1.0	0.9	0.77	0.66	0.5
	IA-W	1.0	0.95	0.93	0.92	0.85
IM5	SS	0.97	0.94	0.92	0.92	0.83
	IA-DCT	1.0	0.94	0.85	0.72	0.57
	IA-W	1.0	0.96	0.95	0.95	0.91
IM6	SS	0.36	0.27	0.22	0.24	0.18
	IA-DCT	1.0	0.85	0.72	0.59	0.46
	IA-W	1.0	0.93	0.88	0.89	0.88
IM7	SS	1.0	0.82	0.78	0.64	0.54
	IA-DCT	1.0	0.95	0.8	0.68	0.53
	IA-W	1.0	0.98	0.95	0.93	0.92
IM8	SS	0.38	0.38	0.38	0.37	0.30
	IA-DCT	1.0	0.92	0.79	0.68	0.52
	IA-W	1.0	0.86	0.84	0.84	0.84

to one-sixteenth of the original size. In all cases, the center portion of the image is kept. The operations here result in a loss of frequency components as well as spatial components. The table clearly shows that the IA-W technique outperforms both the IA-DCT and SS techniques. Although the structure of the IA-DCT scheme is inherently vulnerable to JPEG compression, the global transform of the SS scheme is vulnerable to cropping. This is because the cropping corresponds to convolving the frequency components with a sinc function where the width of the main lobe is inversely proportional to the width of the cropped window size. This will affect all the frequency components of any scheme based on a global transform whereas the IA-DCT and IA-W schemes produce watermarks with local spatial support which are unaffected by the cropping operation. In Table V, *** denotes correlation results close to zero where watermark detection has failed. As predicted, the IA-DCT and IA-W schemes are more robust to cropping than the SS approach. It is important to note that in these experiments, we are making a binary decision on whether a particular watermark exists in a given image. All the results presented here are based on being able to detect one watermark. We extend this work to the detection of multiple watermarks and the issue of capacity in [28].

Robustness to Scaling: Table VI shows the correlation output after scaling where the original image is lowpass filtered

TABLE V
COMPARISON OF ROBUSTNESS TO JPEG COMPRESSION + CROPPING (1/16 ORIGINAL)

Image	Algorithm	quality factor Q				
		crop	crop+Q80	crop+Q60	crop+Q40	crop+Q20
IM1	SS	***	***	***	***	***
	IA-DCT	1.0	0.86	0.73	0.60	0.47
	IA-W	1.0	0.99	0.97	0.94	0.87
IM2	SS	0.50	0.40	0.26	0.26	0.14
	IA-DCT	1.0	0.85	0.83	0.61	0.45
	IA-W	1.0	0.99	0.89	0.83	0.74
IM3	SS	***	***	***	***	***
	IA-DCT	1.0	0.80	0.67	0.53	0.40
	IA-W	1.0	0.94	0.89	0.86	0.85
IM4	SS	0.45	0.4	0.3	0.3	0.3
	IA-DCT	1.0	0.94	0.78	0.67	0.67
	IA-W	1.0	0.96	0.93	0.92	0.92
IM5	SS	0.67	0.78	0.79	0.66	0.57
	IA-DCT	1.0	0.88	0.80	0.66	0.52
	IA-W	1.0	0.97	0.95	0.91	0.89
IM6	SS	***	***	***	***	***
	IA-DCT	1.0	0.83	0.74	0.59	0.47
	IA-W	1.0	0.97	0.92	0.92	0.88
IM7	SS	0.5	0.46	0.45	0.42	0.40
	IA-DCT	1.0	0.96	0.78	0.69	0.004
	IA-W	1.0	0.94	0.93	0.92	0.86
IM8	SS	0.39	0.30	0.20	0.20	0.18
	IA-DCT	1.0	0.91	0.79	0.68	0.54
	IA-W	1.0	0.91	0.88	0.89	0.87

TABLE VI
COMPARISON OF ROBUSTNESS TO SCALING

Scheme	IM1	IM2	IM3	IM4	IM5	IM6	IM7	IM8
SS	0.21	0.12	0.25	0.32	0.46	0.06	0.37	0.25
IA-DCT	0.62	0.78	0.06	0.14	0.3	0.16	0.10	0.22
IA-W	0.93	0.90	0.26	0.86	0.84	0.40	0.84	0.46

using a four-tap filter followed by downsampling by 2 in each direction. The received image is upsampled before the correlation operation is performed. The results vary for each image, but in general, the wavelet scheme outperforms the other two techniques.

Bitstream Watermarking: The DCT-based scheme offers the advantage that if the original data is stored as a JPEG bitstream, the watermark can be inserted directly into the partially decompressed bitstream after entropy decoding and inverse quantization. In contrast, the wavelet-based scheme requires the additional steps of JPEG decoding, wavelet analysis, watermark encoding, wavelet synthesis, and JPEG decoding. One of the concerns in bitstream watermarking using the approaches presented here is JND accuracy. Watermark encoding of a compressed bitstream requires calculating the JND threshold values from a compressed image rather than the original image. We found that the estimate of the JND's based on a compressed image of relatively high quality ($Q \geq 40$) results in a good estimate of the true JND. Another important issue for bitstream watermarking is the additional number of bits required to encode the watermark sequence. Although the addition of the watermark based on perceptual considerations should not affect the *perceptual entropy* of the final image, it could have a significant effect on the signal entropy requiring a greater number of bits to encode using traditional coders. Table VII shows the number of bytes required to encode the original image using baseline JPEG as well as the number of bytes required to encode the IA-DCT watermarked image with the same fixed quality factor Q for both cases. It is clear that the increase in bit rate to encode the watermarked sequence is insignificant where, on average, the increase in bit rate is approximately 2% for a quality factor of 80 and 1% for lower quality factors.

TABLE VII
SIZE OF COMPRESSED ORIGINAL IMAGE VERSUS WATERMARKED IMAGE WITH FIXED QUALITY FACTOR (IN BYTES)

Image	Q80	Q40	Q20
IM1	82289	40537	26087
IA-DCT IM1	83852	41092	26285
IM2	57878	28648	18739
IA-DCT IM2	59164	29500	18987
IM3	65360	33262	21843
IA-DCT IM3	66830	33716	22050

Failures: Although the proposed perceptual schemes as well as the original SS approach are quite robust to many typical image transformations, there are several ways to alter the watermark so that a simple correlation detector is not effective. Since the original image is available for watermark detection, registration of the received, possibly shifted, cropped, or resampled image with the original image, can usually be performed by a simple correlation operation between the two images. There may be instances, however, when image registration becomes difficult so that schemes which are robust to misalignments are important. In our experiments, the IA-DCT scheme as well as the SS scheme are not robust to misalignments while the IA-W scheme does considerably better (see Table VIII). For sample images IM1–IM8, each watermarked image is shifted to the right by one, two, three, and four pixels. Typically a small correlation value (< 0.2) was recovered for the SS and IA-DCT approaches for shifts of one pixel and the values dropped to approximately zero for greater shifts. This is in contrast to the wavelet approach which preserves rather high correlation values in the upper frequency bands. A straightforward way to help address this problem is to include several shifted versions of the received image in the detection process. Subpixel interpolation also poses problems for the SS and IA-DCT techniques. In these experiments we expand all the previous images by 1.6% in each direction by simply removing the last eight rows and columns and expanding the images back to the original size using bilinear interpolation. Watermark detectability is destroyed for all cases

TABLE VIII
IA-W CORRELATION VALUES FOR MISALIGNMENT

column shift	IM1	IM2	IM3	IM4	IM5	IM6	IM7	IM8
1	0.9	0.9	0.84	0.73	0.74	0.80	0.84	0.76
2	0.88	0.88	0.78	0.72	0.74	0.76	0.87	0.75
3	0.87	0.87	0.78	0.72	0.64	0.75	0.89	0.73
4	0.87	0.70	0.75	0.70	0.65	0.70	0.90	0.70

TABLE IX
IA-W CORRELATION VALUES FOR SUBPIXEL INTERPOLATION

Scheme	IM1	IM2	IM3	IM4	IM5	IM6	IM7	IM8
IA-W	0.91	0.90	0.93	0.94	0.91	0.93	0.93	0.89

using the IA-DCT and SS techniques. Only the IA-W scheme survives this operation as shown by the correlation results in Table IX. Since for our applications the original image is available for watermark detection, we can always attempt to estimate the transformations that were applied to the received watermarked image to provide better detection results. As pointed out in [1], however, a collusion attempt, where several versions of the same content with different watermarks are available, will successfully remove the watermarks. Perhaps, the watermarking problem itself needs to be better defined as well as the useful applications before we can determine how robust an algorithm needs to be to be effective.

VI. CONCLUSION

We have introduced image-adaptive watermarking schemes using visual models originally developed for compression applications. In the most general framework, any visual model which provides JND threshold values can be used. Unlike the compression allocation, watermark encoding is not constrained by the amount of side information needed to transmit the perceptual information to the decoder. Therefore, the perceptual thresholds, as given by the JND's obtained from the visual models, can be fully utilized. Two perceptual schemes have been proposed: the IA-DCT and IA-W approaches. The IA-DCT algorithm offers the advantage of being able to watermark partially decompressed JPEG bitstreams. The results show that the DCT framework of the IA-DCT scheme is quite robust to JPEG compression as well as other types of common image transformations. Although the IA-W scheme is based on a much simpler visual model which only takes into account frequency sensitivity, the multiresolution structure of the watermark and the watermark detection scheme results in a very robust scheme. In general, the IA-W scheme yields the overall best results. More sophisticated visual models in the wavelet framework should further improve the current results. Visual models providing JND values which take into account temporal masking as well as spatial masking could be used to extend these results to video watermarking.

ACKNOWLEDGMENT

The authors thank L. O'Gorman for introducing them to the problem and for many fruitful discussions. They would also like to thank B. Safranek for providing some of the code for the experiments.

REFERENCES

[1] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," NEC Research Institute Tech. Rep. 95-10, 1995.
[2] J. Brassil, S. Low, N. Maxemchuk, and L. O'Gorman, "Electronic marking and identification techniques to discourage document copying," in *Proc. Infocom'94*, pp. 1278–1287.
[3] R. Schyndel, A. Tirkel, and C. Osborne, "A digital watermark," in *IEEE Proc. Int. Conf. Image Processing*, 1994, vol. 2, pp. 86–90.
[4] G. Caronni, "Assuring ownership rights for digital images," in *Proc. Reliable IT Systems, VIS'95*, 1995.
[5] W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," MIT Media Lab Tech. Rep., 1994.
[6] K. Matsui and K. Tanaka, "Video-steganography," in *IMA Intellectual Property Project Proc.*, 1994.
[7] F. Goffin, J. F. Delaigle, C. De Vleeschouwer, B. Macq, and J. J. Quisquater, "A low cost perceptive digital picture watermarking method," in *Proc. SPIE Electronic Imaging '97: Storage and Retrieval of Image and Video Databases V*, Feb., vol. 3022, pp. 264–277.
[8] I. Pitas, "A method for signature casting on digital images," in *IEEE Proc. Int. Conf. Image Processing*, 1996, vol. 3, pp. 215–218.
[9] G. C. Langelaar, J. C. A. van der Lubbe, and R. L. Lagendijk, "Robust labeling methods for copy protection of images," in *Proc. SPIE Electronic Imaging '97: Storage and Retrieval of Image and Video Databases V*, Feb., vol. 3022, pp. 298–309.
[10] R. B. Wolfgang and E. J. Delp, "A watermark for digital images," in *IEEE Proc. Int. Conf. Image Processing*, 1996, vol. 3, pp. 219–222.
[11] S. Craver, N. Memon, B.-L. Yeo, and M. Yeung, "Can invisible watermarks resolve rightful ownerships?," in *Proc. SPIE Electronic Imaging '97: Storage and Retrieval of Image and Video Databases*, vol. 3022, pp. 310–321.
[12] E. Koch and J. Zhao, "Toward robust and hidden image copyright labeling," presented at the Nonlinear Signal Processing Workshop, Thessaloniki, Greece, 1995.
[13] A. G. Bors and I. Pitas, "Embedding parametric digital signatures in images," in *Proc. Eur. Conf. Signal Processing, EUSIPCO96*, Trieste, Italy.
[14] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Transparent robust image watermarking," in *IEEE Proc. Int. Conf. Image Processing*, 1996, vol. 3, pp. 211–214.
[15] J. O. Ruanaidh, W. J. Dowling, and F. M. Boland, "Phase watermarking of digital images," in *IEEE Proc. Int. Conf. Image Processing*, 1996, vol. 3, pp. 239–242.
[16] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for images, audio and video," in *IEEE Proc. Int. Conf. Image Processing*, 1996, vol. 3, pp. 243–246.
[17] C. I. Podilchuk and W. Zeng, "Digital image watermarking using visual models," in *Proc. SPIE Human Vision and Electronic Imaging II*, Feb. 1997, vol. 3016, pp. 100–111.
[18] ——, "Perceptual watermarking of still images," in *Proc. IEEE Workshop on Multimedia Signal Processing*, June 1997.
[19] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.
[20] H. A. Peterson, A. J. Ahumada, Jr., and A. B. Watson, "Improved detection model for DCT coefficient quantization," in *Proc. SPIE Conf. Human Vision, Visual Processing, and Digital Display IV*, Feb. 1993, vol. 1913, pp. 191–201.
[21] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE Conf. Human Vision, Visual Processing, and Digital Display IV*, Feb. 1993, vol. 1913, pp. 202–216.
[22] R. J. Safranek, "A comparison of the coding efficiency of perceptual models," in *Proc. SPIE Conf. Human Vision, Visual Processing, and Digital Display VI*, 1995.
[23] ——, "Perceptually based prequantization for image compression," in *Proc. SPIE Conf. Human Vision, Visual Processing, and Digital Display V*, 1994.
[24] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image-dependent quantization and post-quantization data compression," in *IEEE Proc. ICASSP*, 1989.
[25] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visual thresholds for wavelet quantization error," in *Proc. SPIE Human Vision and Electronic Imaging*, 1996, vol. 2657, pp. 381–392.

[26] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using the wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, 1992.
[27] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968.
[28] S. Servetto and C. Podilchuk, "Capacity issues for digital watermarking," in *IEEE Int. Conf. Image Processing '98*, submitted.

Christine I. Podilchuk (S'83–M'91) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Rutgers University, Piscataway, NJ, in 1984, 1986, and 1988, respectively.

Currently, she is working in the Visual Communications Research Department at Bell Labs, Lucent Technologies, Murray Hill, NJ. Her research interests are in the general area of image processing and computer vision. Her previous work includes video compression and transmission, face recognition, and digital watermarking of images and video.

Dr. Podilchuk is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.

Wenjun Zeng was born in Longyan, China. He received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 1990, the M.S. degree in electrical engineering from University of Notre Dame, Notre Dame, IN, in 1993, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in June 1997.

From 1990 to 1991, he worked as a Research Staffer at the Department of Electronic Engineering, Tsinghua University, where he was responsible for developing a fast image/video processing system. He was with Matsushita Information Technology Laboratory, Panasonic Technologies Inc., Princeton, NJ, in 1995 and with Multimedia Communication Laboratory, Lucent Technologies, Bell Laboratories, Murray Hill, NJ, in 1996. Since June of 1997, he has been a Member of Technical Staff in the Digital Video Department of Sharp Laboratories of America, Inc., Camas, WA. His research interests include image and video signal processing, resilient video transmission, digital watermarking, and data hiding.